



大连理工大学

信息检索研究室

Information Retrieval Laboratory of DUT

语言计算 顶天立地

林鸿飞

2012年04月13日 江西婺源

- 中文信息处理历程
- 语言计算的普适性
- 计算思维与语言计算
- 语言计算应用举例
- 结束语

- **文字信息处理（过去的三十年）**

主要解决了汉字的输入、输出、识别等，让中文走进计算机。

- **语言信息处理（下一个三十年）**

顶天：凝练中文特色，提高核心竞争力

背景：经济建设 文化战略 信息安全

立地：走入千家万户，服务人文社会学科

背景：海量数据、语言信息、社会需求

- **交叉特点：**

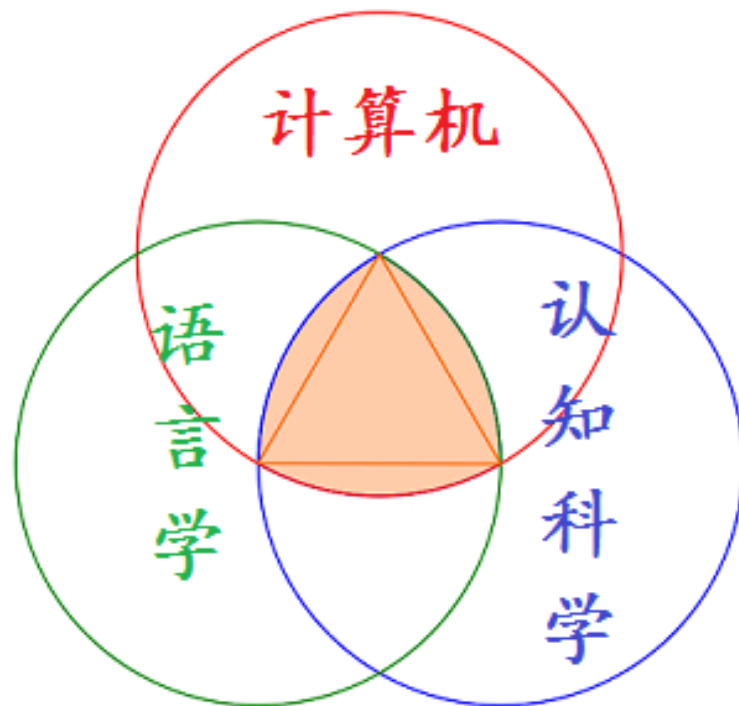
**计算机、语言学、认知科学等学科交叉
(语言计算)**

- **研究方法：**

**从语言学视角加以谋划
以数学形式加以描述
以工程方法加以实现**

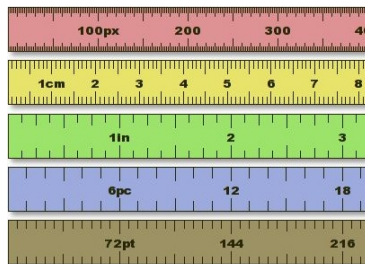
- **工具属性：**

对于相关学科的支撑作用



语言的功能：

- 交际
- 思维工具
- 文化载体



语言计算的作用：

- 像显微镜、天平、量尺等作为物理世界的研究工具
- 使分词、语法语义分析、语言模型等成为以语言文字为对象的人文社会科学的基本研究工具

- **计算思维：周以真（Jeannette M.Wing）**
- **语言计算是体现在人文社会科学领域计算思维的最好的载体和工具。**
- **一门科学只有在成功运用了数学时才算达到了完善的地步。 ---马克思**
- **以语言文字为对象的人文社会科学运用了语言计算可以获得更大的发展动力。**

- **资源**（通用资源、特定领域资源）
- **平台**（核心软件、基础平台）
- **工具**（面向领域的程序）
- **案例**（实施的项目库）

普及语言计算的途径



群体智慧



海量数据



云计算

机器学习



- 为语言计算找到应用的场所和背景，促进相关学科的发展。
- 从应用领域中凝练科学问题，提升语言计算的研究水平。
- 吸引更多具有交叉学科背景的研究人员加入自然语言处理的行列中，扩大中文信息学会的影响。
- 提高学校和社会上对于语言计算的认识，积极影响和参与国家信息、文化、教育等重大决策。

科技管理：文献计量与知识图谱

- 查询构造与命名实体识别
- 基于内容的计量（传统的共词、共引分析）
- 引用分析（涉及到正向负向引用）

国家社科基金项目

英汉典籍辅助翻译：

- 术语识别与双语对齐
- 不同译本分析
- 不同译者分析

国家社科基金项目

电子商务：

- 网络营销
- 产品推荐

《情感词汇本体》 《情感语料库》 《情感常识知识库》

国家自然科学基金项目

互联网侦查学：

- 海量数据采集、清洗、索引、存储、检索、统计等
- 社会关系网络分析
- 关联挖掘
- 机会发现

从平常中发现不平常，从不平常中发现平常

公安部科技创新项目



- **语言计算 方兴未艾**
- **服务社会 价值凸显**
- **群体智慧 人机和谐**
- **凝练特色 扩大影响**



谢谢！